

KMU-Ratingmodelle und Ratingqualität: Auswirkungen der Ratingarchitektur auf die ex-ante Risikoklassifikation von KMU-Kreditkontrakten



Beat Bernet, Simone Westerfeld

Überblick

- Die Bonitätsklassifikation einer KMU als Nachfragerin eines Kredites bei einer Bank bestimmt wesentlich die Kreditkonditionen. Aus Sicht der KMU ist daher die Frage von Bedeutung, inwiefern die Ausgestaltung bankinterner Ratingmodelle bei identischen Ausgangsdaten das Ergebnis des Ratingprozesses bestimmt.
- Ziel dieses Beitrages ist es, die Auswirkungen unterschiedlich gestalteter Architekturen von Ratingmodellen auf die ex-ante Risikoklassifikation von KMU zu analysieren. Dazu wird ein für den schweizerischen KMU-Markt repräsentatives Portfolio von über 400 Kreditnehmern mit drei typischen, realen Ratingmodellen von Banken bewertet.
- Die Untersuchung bestätigt, dass differierende Ratingarchitekturen bei gleichem Informationsinput pro Kreditnehmer zu signifikant unterschiedlichen Ratings führen. Gleichzeitig muss die auf empirische Arbeiten gestützte Hypothese verworfen werden, dass der Einbezug qualitativer Informationen in der Tendenz zu höheren Bonitätseinschätzungen führt als solche, die ausschliesslich auf quantitativen Informationen basieren.

Keywords Banks · rating models · SME financing · ex-ante risk classification · model architecture

Eingegangen: 7. Juli 2007

JEL: G21, G32

Prof. Dr. Beat Bernet (✉)

Schweizerisches Institut für Banken und Finanzen (s/bf), Universität St. Gallen (HSG), Rosenbergstrasse 52, 9000 St. Gallen, Email: beat.bernet@unisg.ch, URL: www.sbf.unisg.ch.

Dr. Simone Westerfeld (✉)

Schweizerisches Institut für Banken und Finanzen (s/bf), Universität St. Gallen (HSG), Rosenbergstrasse 52, 9000 St. Gallen, Email: simone.westerfeld@unisg.ch, URL: www.sbf.unisg.ch.

A. Einführung

Mit der Einführung des neuen Basler Accords hat die Diskussion bankinterner Ratingmodelle in Wissenschaft und Praxis an Bedeutung und Intensität gewonnen. Zwar spezifiziert der Basler Ausschuss für Bankenaufsicht („Basler Ausschuss“) in seinem Grundlagenpapier konzeptionelle Anforderungen, denen interne Ratingmodelle zu genügen haben, um die Akzeptanz der Aufsichtsbehörden zu erreichen (Basler Ausschuss, 2006). Innerhalb dieser Leitplanken geniessen Banken und Finanzinstitute aber bei der Ausgestaltung ihrer Ratingmodelle einen grossen Freiraum. Als Konsequenz daraus werden sich auch bei Anwendung des Standardansatzes in den kommenden Jahren bankinterne Ratingmodelle zur Beurteilung der Bonität von KMU-Kreditnehmern hinsichtlich Ratingphilosophie, Systemarchitektur oder Parametrisierung weiterhin voneinander unterscheiden.

Aus der Sicht einer KMU als aktuelle oder potentielle Kreditnachfragerin stellt sich damit die Frage, ob und wie diese unterschiedliche Ausgestaltung interner Ratingmodelle verschiedener Kreditanbieter bei grundsätzlich identischen Inputdaten die Bonitätsklassifikationen und damit letztlich die Kreditkonditionen beeinflusst. Mit den Auswirkungen auf die Bonitätsklassifikation befasst sich der folgende Beitrag. Während externe Ratingmodelle grosser Agenturen aus unterschiedlichen Erkenntnisperspektiven detailliert analysiert wurden (Cantor/Packer, 1997; Poon, 2003; Löffler, 2004), gibt es bisher kaum empirische Untersuchungen, die sich mit den Auswirkungen der Architektur interner Ratingmodelle auf die ex-ante Risikoquantifizierung von Kreditkontrakten befassen. Die Entscheidung einer Bank, einer KMU einen Kredit zu gewähren sowie die damit verbundenen Konditionen wie Volumen, Laufzeit oder Preis sind nicht nur für den einzelnen Kreditnehmer, sondern für die ganze Volkswirtschaft von Bedeutung. Die Effizienz der Antragsselektion beeinflusst die Kreditversorgung der Wirtschaft, die langfristige Abschreibungsquote und damit die Ertragskraft der Banken sowie, angesichts der Bedeutung dieser Kreditkontrakte für die Finanzbranche, die Stabilität eines Finanzsystems.

Ausgehend von einer kurzen Darstellung der Bedeutung unterschiedlicher Ratingmodelle und einer Analyse von Beiträgen zur Ratingarchitektur in Kapitel A, formulieren wir zu Beginn des Kapitels B zwei Hypothesen und skizzieren anschliessend drei Ratingmodelle unterschiedlicher Architekturklassen, die stellvertretend für die in der Bankenpraxis heute eingesetzten Modelle als Basis der empirischen Untersuchung dienen. Die Beschreibung der Datenbasis und der Vorgehensweise schliessen den Abschnitt ab. Kapitel C stellt die empirischen Ergebnisse der drei Ratingmodelle im Vergleich dar. Kapitel D fasst die Ergebnisse zusammen und leitet Schlussfolgerungen ab.

In Anlehnung an die Definition des Basler Ausschusses verstehen wir unter dem Begriff **Ratingsystem** alle Methoden, Prozesse, Kontrollen, Datenerhebungen und die dazu eingerichteten und genutzten Informationsverarbeitungssysteme, die der Zuweisung bankinterner Ratingklassifizierungen und der Ableitung von Ausfall- und Verlustschätzungen dienen (Basler Ausschuss, 2006). Der im Folgenden verwendete Begriff des **Ratingmodells** beschreibt als Teilmenge des Ratingsystems das in der Bank zur Erhebung und Quantifizierung der Bonitätsklassifikation bzw. der Zuordnung eines konkreten Ratingurteils für einen Kreditnehmer eingesetzte Instrument. Diesem Urteil wird typischerweise eine durchschnittliche Ausfallwahrscheinlichkeiten („probability of default“ bzw. PD)

attribuiert. Die **Architektur eines Ratingmodells** umfasst wiederum die einzelnen Elemente des Ratingmodells sowie die Ausgestaltung ihrer Beziehung zueinander. Desweiteren definieren wir **Ratingklassen** als Teil der Ratingarchitektur, mit Hilfe derer die Einteilung der Bonität in verschiedene Stufen erfolgt. Die **Ratingklassifikation** beschreibt schliesslich den konkreten Output bei Anwendung des Ratingmodells in Form der Zuweisung einer Ratingklasse zur Bonitätseinschätzung.

Ein Rating beschreibt damit immer ein ex-ante quantifiziertes Risiko, dass das klassifizierte Objekt seinen Verpflichtungen nicht nachkommen kann bzw. wird. Ein Blick in die Ratingpraxis der Banken zeigt, dass PDs oft die wichtigste Entscheidungsgrundlage für die Gewährung bzw. Ablehnung von Krediten, aber auch für die Berechnung von Risikoprämien und Kreditlimiten sind.

I. Aspekte der Modellarchitektur

Literatur wie Praxis unterscheiden bezüglich der verwendeten Verfahren zwischen quantitativen und qualitativen sowie hybriden Ratingmodellen (Füser, 2001; Oelerich, 2004). In qualitativen Ratingmodellen bewerten Experten die definierten Ratingkriterien. Die erfassten Merkmale lassen sich meist nur subjektiv bestimmen, erfassen, gewichten, verknüpfen und bewerten, selbst wenn sie quantitativ bestimmt werden können wie beispielsweise im Fall von Marktanteilen. Quantitative Modelle basieren auf standardisierten, objektiven und mathematisch-statistischen Verfahren. Meist enthalten aber auch diese Modelle subjektive Elemente wie etwa die Definition von Regeln zur Verknüpfung von Elementen. So wie qualitative Modelle auch quantitative Daten verarbeiten, können quantitative Modelle auch qualitative Daten verwenden – massgeblich für die Zuordnung ist nicht die Art der Information, sondern die Art der **Informationsverarbeitung** (Eigermann, 2002; Oelerich, 2004). Qualitative Verfahren basieren auf Punktbewertungsverfahren, die aus Expertenurteilen, aus Systemen künstlicher Intelligenz oder anderen subjektiven Beurteilungsverfahren abgeleitet werden. Quantitative Modelle basieren auf statistischen Verfahren (wie Regressionsmodellen oder Diskriminanzanalysen) oder sonstigen mathematischen Verfahren. Ratingmodelle, die qualitative und quantitative Elemente simultan oder sequentiell verwenden, werden als hybride Modelle bezeichnet.

Die Modellarchitektur folgt im internationalen Umfeld offensichtlich gemeinsamen Prinzipien. In einem ausführlichen Diskussionspapier befasste sich die ‚Models Task Force‘ des Basler Ausschusses mit einer empirischen Analyse und Klassifizierung interner Ratingmodelle von Banken (Basler Ausschuss, 2000a). Die Untersuchung zeigte, dass zwar im Untersuchungszeitpunkt kein einheitlicher Standard für Design und Einsatz interner Ratingmodelle bestand, dass aber doch eine Anzahl allen Systemen gemeinsamer Elemente festgestellt werden konnte. Dazu gehören die Wahl der Risikoindikatoren, die Priorisierung des Schuldnerratings über dasjenige der Transaktion oder Aspekte der Strukturierung von Ratingprozessen. Eigermann (2002) untersucht hybride Ratingsysteme deutscher Banken. Bernet/Westerfeld (2007) zeigen, dass bankinterne Ratingmodelle immer anhand so genannter Architekturelemente sowie einer Reihe von Metaentscheidungen zu Ratingphilosophie, Ratingobjekt, Mapping und Validierung charakterisiert werden können. Krahen/Weber (2001) entwickelten darüber hinaus dreizehn Prinzipien, die als Qualitätsstandard zur Entwicklung von adäquaten bankinternen Ratingmodellen

für das Gegenparteirisiko verstanden werden können. Zu nennen sind hier u.a. allgemeine Erfordernisse wie ein angemessener Umfang und Vollständigkeit des Ratings, um alle vergangenen, aktuellen und zukünftigen Kunden raten zu können und eine vollständige Datenbank zu unterhalten, sowie eine adäquate Komplexität des Ratingmodells, um Kunden verschiedenster Risikoklassen differenziert beurteilen zu können. Ähnlich argumentiert die Bank for International Settlement in einem Konsultationspapier, das die kumulativ zu erfüllenden Anforderungen beschreibt, denen Ratingmodelle zu genügen haben, um das Anerkennungsverfahren nationaler Aufsichtsinstanzen zu bestehen (Basler Ausschuss, 2006).

Beiträge von Carey (2002), Treacy/Carey (2000) und English/Nelson (1998) beschreiben und analysieren interne Ratingmodelle grosser amerikanischer Banken. Die Untersuchung von Treacy/Carey zeigt nicht nur die Architekturelemente von Ratingmodellen dieser Banken auf, sondern weist auch auf die vielfältigen Probleme hin, die aus inkonsistenten Modellarchitekturen entstehen können. Ähnliche Untersuchungen für schwedische Banken präsentieren Jacobson/Lindé/Roszbach (2006) sowie Nakamura/Roszbach (2005). Eine weitere Studie von Hornik et al. (2005) basiert auf einem Datensample der österreichischen Nationalbank mit sich teilweise überschneidenden Kundensamples von 27 Banken. In der Studie wird ein Ansatz entwickelt, um Multi-Rater Informationen zur Validierung von Ratingmodellen zu verwenden, insbesondere für Situationen, in denen keine historischen Ausfalldaten vorliegen.

II. Bedeutung unterschiedlicher Ratingarchitekturen

Banken als die wichtigsten Kreditgeber für KMU unterscheiden sich stark hinsichtlich ihrer Selektionskriterien bei der Behandlung von Kreditanträgen bzw. bei deren risikoadjustierter Bewertung. Tatsächlich zeigt ein Blick in die KMU-Ratingpraxis jedem Kreditnehmer, der mit verschiedenen Banken Kreditbeziehungen unterhält, dass bei gleicher Ausgangslage die individuelle Risikoklassifizierung durch verschiedene Banken meist unterschiedlich ausfällt. Unterschiedliche Zuordnung von Ratings durch unterschiedliche Modelle bei identischer Datenlage wirken sich direkt auf die Kapitalkosten der kreditsuchenden KMU aus. Umgekehrt haben unterschiedliche Modellarchitekturen aber auch Konsequenzen für die Wettbewerbsposition der Kreditanbieter, da ein ‚falsches‘ Rating – bei sonst gleichen Konditionen – zu adversen Selektionseffekten bzw. zu Marktanteilsverlusten und Ertragsausfällen führt.

In der jüngeren Vergangenheit sind immer mehr Banken dazu übergegangen, den kreditsuchenden Unternehmungen die Kernelemente ihres Ratingprozesses offen zu legen. Das eingesetzte Ratingmodell wird dabei aber bestenfalls in seinen wichtigsten Parametern beschrieben (z.B. hinsichtlich der verwendeten qualitativen und quantitativen Inputfaktoren), die Beziehung der Elemente zueinander als Teil der Modellarchitektur wird jedoch kaum kommuniziert. Das damit verbundene Wissen, das letztlich zu dem Rating führen soll, das dem tatsächlichen Ausfallverhalten am nächsten kommt, wird als wichtiger Differenzierungsfaktor im Wettbewerb betrachtet. Für den Kreditnachfrager wie für die Konkurrenten bleibt das Modell damit eine ‚Black Box‘, zu der bestenfalls Input und Output bekannt ist, nicht aber die Modellarchitektur und die vom Input zum Rating führenden Transformationsgleichungen.

Die Frage, ob und in welchem Ausmass die Architektur des Ratingmodells die Risikoklassifizierung eines Schuldners beeinflusst, ist deshalb sowohl aus makro- wie aus mikroökonomischer Sicht von Interesse.

Darüber hinaus ist die Untersuchung bankinterner Ratingmodelle relevant, da bisher kaum empirische Untersuchungen zur ex-ante Risikoquantifizierung dieser Modelle gemacht wurden und sich die Erkenntnisse aus der Analyse externer Ratingmodelle nicht auf interne Modelle übertragen lassen. Dies deshalb, da sich externe Ratings üblicherweise nicht auf das Ratingobjekt KMU, sondern auf grosse, börsennotierte Unternehmen beziehen und der Markt der externen Ratings für KMU noch in den Anfängen seiner Entwicklung steckt (Kley, 2003). Ausserdem basieren externe Ratings auf einer Architektur, die in der Methode weniger auf mathematisch-statistische Verfahren als vielmehr auf Expertenwissen zurückgreift und einen Ansatz des ‚through-the-cycle‘ Ratings verwendet, der gemäss Altman/Rijken (2004) zu stabileren Ratingergebnissen führt als andere, auf einen gewissen Zeithorizont ausgerichtete Modelle. Zusätzlich fliessen in die Bewertung von KMU durch eine Bank zahlreiche private Informationen aus der engen Beziehung zwischen den Akteuren ein, die das interne Rating vom externen Rating unterscheiden (Mommel/Schmieder/Stein 2006). Wir testen die Hypothese, dass unterschiedliche Ratingarchitekturen zu unterschiedlichen Risikoklassifikationen führen, in Abschnitt C.I.

III. Bedeutung qualitativer und quantitativer Ratingkriterien

Ratingmodelle können, wie ausgeführt, auf qualitativen oder quantitativen Kriterien bzw. auf einer entsprechenden Kombination aufgebaut sein. Auch hier ist die Frage von Interesse, ob bei objektiv gleicher Ausgangslage eines Schuldners die Anwendung qualitativer oder quantitativer Kriterien zu einem anderen Klassifizierungsergebnis führt bzw. wie der Mix der beiden Kriterienkategorien das Ergebnis beeinflusst. Bisherige empirische Untersuchungen legen die Vermutung nahe, dass qualitative Kriterien systematisch besser bewertet werden als quantitative Kriterien und damit Ratingmodelle mit einem höheren Anteil bzw. Gewicht der qualitativen Kriterien zu einer besseren Ratingklassifikation führen. So zeigen etwa Brunner/Krahn/Weber (2000) in einer breiten vergleichenden Untersuchung der Ratingmodelle dreier deutscher Banken, dass qualitative Kriterien systematisch und signifikant besser bewertet werden als quantitative Kriterien, und dass sie über die Risikoklassen hinweg eine deutlich geringere Varianz zeigen. Die Analyse unterscheidet sich von der im Folgenden präsentierten empirischen Untersuchung insofern, als dass Ratingdaten von drei deutschen Banken verwendet wurden. Die Banken haben je ein zufälliges Sample von 200 Kreditnehmern aus ihrem Kreditportfolio gezogen, die anschliessend das Datensample der Untersuchung bildeten. Die Analyse wurde anhand einer auf nur vier Ratingklassen gemappten Skala vorgenommen, so dass keine Aussagen darüber gemacht werden können, ob eine heute in der Ratingpraxis meist anzutreffende (und von Basel II auch verlangte) Verfeinerung auf zehn oder mehr Ratingklassen zu gleichen Ergebnissen führen würde. Zu einem ähnlichen Ergebnis kamen bereits früher Weber/Krahn/Vossmann (1999), die nicht nur zeigten, dass qualitative Teilratings als Element einer Risikoklassifizierung für den gleichen Schuldner systematisch höher ausfielen als die entsprechenden quantitativen Teilratings, sondern auch nachwiesen, dass

qualitative Teilratings über die Zeit weniger stark schwanken (d.h. eine geringere Migration aufweisen) als quantitative Teilratings.

Wir überprüfen in Abschnitt C.II die Hypothese, dass eine Risikoklassifizierung unter Verwendung qualitativer Kriterien zu einer systematisch höheren Bonitätseinschätzung im Vergleich zur Risikoklassifizierung ausschliesslich auf der Basis von quantitativen Kriterien führt.

B. Empirische Analyse

Wenn eine unterschiedliche Architektur von KMU-Ratingmodellen zu unterschiedlichen Ratingklassifizierungen führen kann, so ist angesichts der Bedeutung des Ratingurteils für die Kapitalallokation, die Wettbewerbsposition der Kreditanbieter und die Kapitalkosten bzw. den Kapitalzugang für die Kreditnachfrager die Frage von Bedeutung, ob und wie das Rating durch die Modellarchitektur beeinflusst wird. Mit dem ersten Teil dieser Fragestellung befasst sich die nachstehende empirische Untersuchung. Sie bewertet ein schweizerisches KMU- Kreditportfolio durch Ratingmodelle mit unterschiedlichen Architekturen und versucht festzustellen, ob diese unterschiedliche Ausgestaltung der Modellarchitektur bei grundsätzlich identischen Inputdaten eines Kreditnehmers zu unterschiedlichen Bonitätsklassifikationen und damit zu unterschiedlichen Kreditkonditionen führt.

Wir gehen dabei auf Basis der in den obenstehenden Abschnitten angeführten Literatur von folgenden Hypothesen aus, die wir mit Hilfe des Samples für den KMU-Kreditmarkt der Schweiz überprüfen:

- H₁:** Unterschiedliche Ratingarchitektur führt für den Kreditnehmer c.p. zu einem signifikant unterschiedlichen Rating.
- H₂:** Der Einbezug qualitativer Informationen führt zu einer signifikant höheren Bonitätseinschätzung im Vergleich zur Ratingevaluation auf der Basis rein quantitativer Informationen.

Um diese Hypothesen empirisch zu untersuchen, spezifizieren wir im Folgenden drei für den schweizerischen Bankenmarkt typische Ratingmodelle. In den folgenden Abschnitten beschreiben wir die Ratingmodelle, die Datenbasis sowie das gewählte Vorgehen.

I. Charakterisierung der Ratingmodelle

In einem ersten Schritt wurden die bankinternen Ratingsysteme der schweizerischen Kantonal-, Regional- und Raiffeisenbanken analysiert und systematisiert: Die meisten der in der schweizerischen Bankpraxis verwendeten bankinternen Modelle sind der Klasse der statistischen, d.h. schwergewichtig quantitativen Modellen zuzuordnen. Dabei werden mehrheitlich Modelle eingesetzt, die sowohl qualitative wie quantitative Daten verarbeiten. Ein deutlich geringerer Teil von Banken erarbeitet Ratingklassifizierung ausschliesslich auf der Basis von Bilanzdaten, d.h. ohne Einbezug qualitativer Kriterien.

Aus dieser Untersuchung identifizierten wir drei typische Ratingmodelle, anhand derer alle im Markt vorgefundenen Ratingmodelle klassifiziert werden konnten. Zwei Modelle

sind statistischer Ausprägung und der Gruppe der hybriden Modelle zuzuordnen; sie unterscheiden sich primär durch den Anteil quantitativer bzw. qualitativer Inputfaktoren. Das dritte Modell basiert ausschliesslich auf Bilanzdaten und repräsentiert damit ein rein quantitatives Ratingmodell. Diese drei Modelle sind für die in der schweizerischen Bankpraxis verwendeten Ratingmodelle typisch. Rein qualitative Modelle spielen für die schweizerische KMU-Ratingpraxis heute keine nennenswerte Rolle mehr.

Den im Folgenden vorgestellten Modellen ist gemeinsam, dass es sich um Point-in-Time Ratings mit einem 1-Jahreshorizont handelt.¹ Alle drei Modelle basieren auf einem Scoring-Prozess: Ein Kriterium i wird mit einer vorgegebenen Transformationsfunktion $\{v_i\}$ bewertet; anschliessend wird, basierend auf einer Aggregierungsregel, dem individuellen Kriterium ein Gewicht k_i zugeordnet; die Ergebnisse werden schliesslich nach einer algebraischen Regel zu einem generellen Score $v(a)$ zusammengefasst. Der so ermittelte Scorewert bestimmt die Zuteilung zu einer Ratingklasse.²

$$(1) \quad v(a) = \sum_i k_i v_i(a_i)$$

Während die Architekturelemente der drei Modelle grundsätzlich identisch sind, unterscheiden sie sich vor allem im Hinblick auf die verwendeten Kriterien, der Transformation in Scores und deren Gewichtung. Alle drei Modelle sind real existierende Modelle, die in der Ratingpraxis seit Jahren verwendet werden.

Bei **Modell 1** handelt es sich um das Ratingmodell einer Bank, das für die Bewertung von nicht börsennotierten Unternehmen des KMU-Segments eingesetzt wird. Es basiert auf sieben Finanzkennzahlen (quantitative Faktoren aus den Bereichen Liquidität, Finanzierung, Vermögenslage, Rentabilität, Umschlag) und elf qualitativen Faktoren. Die Ermittlung der Gewichtungen der quantitativen und qualitativen Faktoren wird jeweils mit einem statistischen Ansatz vorgenommen: Zur Ermittlung der Kriterien und deren Gewichtung wird zunächst – für quantitative und qualitative Faktoren getrennt – eine univariate Analyse durchgeführt und anhand der Receiver Operating Characteristic (ROC) – einem Mass für die Fähigkeit eines Kriteriums, zwischen in der Zukunft insolventen bzw. solventen Firmen zu unterscheiden – Kennzahlen mit einem akzeptablen Wert für die Trennkraft im Modell selektiert.³ Schliesslich wird in einem iterativen multivariaten Analyseverfahren (Logitmodell) aus den vorselektierten Kriterien diejenige Kennzahlenkombination gesucht, welche die Bonität eines Kunden möglichst gut erklärt. Beide Teilratings liefern einen Scorewert, der wiederum mit einem neuen Logitmodell relative Gewichte zugewiesen bekommt. Der Anteil der qualitativen Faktoren am Gesamtrating beträgt in diesem Modell ca. 30%. Die Ausfallwahrscheinlichkeiten, die als Grundlage des Mappings gebraucht werden, wurden aufgrund historischer Zahlungsausfälle geschätzt und das Rating Tool entsprechend kalibriert. Es kommt eine Ratingskala mit 10 Klassen für ‚performing assets‘ zum Einsatz.

Das **Modell 2** ist ebenfalls ein „Hybrid“; es unterscheidet sich jedoch bezüglich Modellentwicklung, verwendeter Merkmale, Architektur und Kalibrierung substantiell von Modell 1: Die Merkmale wurden sowohl aufgrund ihrer Trennfähigkeit als auch unter Berücksichtigung der Plausibilität und Kommunizierbarkeit des Modells in einem mehrstufigen Verfahren selektiert. Das Gesamtmodell besteht aus einem statistisch optimierten Teilmodell zur Früherkennung der ausfallgefährdeten Schuldner, das auf quantitativen

Kriterien beruht, und einem Expertensystem, welches die Differenzierung zwischen den Schuldnern im oberen Bereich des Bonitätsspektrums anhand von qualitativen Kriterien vornimmt. Um eine adäquate Gewichtung der ökonomisch relevanten Bereiche zu gewährleisten, wurde jede der insgesamt sieben quantitativen Kennzahlen zuerst einer logistischen Transformation unterzogen. Zusammen mit zwei weiteren quantifizierbaren Attributen wurden die derart transformierten Kennzahlen mittels multivariater Optimierung zu einer Scoringfunktion kombiniert. Parallel hierzu wurde ein Expertensystem entwickelt, welches sieben qualitative Faktoren zu einem „Expertenscore« kombiniert und – mittels Identifikation von Risikofaktoren sowie der Beurteilung von Profitabilität und Rentabilität – eine optimale Differenzierung im oberen Bereich des Bonitätsspektrums gewährleisten soll.

Die Verrechnung der Resultate des auf quantitativen Kriterien beruhenden Teilmodells und des Expertensystems erfolgt nach folgender Logik: Zunächst wird anhand des auf der Basis von quantitativen Kriterien ermittelten Scores ein Ausgangsrating ermittelt, welches über den Einfluss des „Expertenscores« entscheidet. Um die Leistungsfähigkeit des statistischen Teilmodells nicht zu kompromittieren, ist für Schuldner mit schwachem Ausgangsrating ausschliesslich dieses ausschlaggebend, d. h. im unteren Bereich der Rating-skala haben die qualitativen Faktoren keinen Einfluss auf das Rating (Gewicht = 0%). Erreicht andererseits das Ausgangsrating einen hinreichend guten Wert, so erfolgt die Primäreinteilung in eine feste Ratingklasse und die definitive Zuweisung des Ratings aufgrund des Scores des Expertensystems, wobei letzteres in diesem Bereich etwa 30% Gewicht hat. Für mittlere Ausprägungen des auf der Basis von quantitativen Kriterien ermittelten Scores wird dieser als Ausgangspunkt verwendet und die definitive Einteilung ebenfalls durch das Expertensystem vorgenommen, wobei dessen Einfluss mit zunehmender Güte des Ausgangswertes von 0% bis ca. 30% anwächst.

Das Modell verwendet grundsätzlich 10 Ratingklassen für „performing assets“, wobei die besten zwei Klassen den Schuldnern des KMU Segmentes nicht zur Verfügung stehen. Eine Primäreinteilung durch das statistische Teilmodell in die Klassen 10 bis 8 ist definitiv; die bestmögliche Primäreinteilung erfolgt in die Klasse 5, wobei in diesem Fall die quantitativen Faktoren eine Endeinteilung in die Klassen 7 bis 3 bewirken können.

Bei **Modell 3** handelt es sich um ein Modell, das ausschliesslich auf Jahresabschlussdaten basiert. Qualitative Faktoren fliessen nicht in das Ratingergebnis ein, jedoch werden Brancheninformationen wie z.B. unterschiedliche durchschnittliche Ausfallerwartungen pro Branche berücksichtigt. Es werden spezifische finanzielle Kennzahlen aus den Jahresabschlüssen der Unternehmen verwendet, welche geeignet sind, die langfristige Performance der Unternehmung auch für ‚out-of-the-sample‘ Firmen zu erfassen. Die Kennzahlen stammen aus den folgenden Gruppen von Risikofaktoren: Profitabilität, Fremdfinanzierungsgrad, Schuldendeckung, Wachstum, Liquidität, Aktivitätskennzahlen und Grösse der Unternehmung. Das Modell berücksichtigt im Design die Tatsache, dass der Zusammenhang zwischen einer Finanzkennzahl und der Ausfallwahrscheinlichkeit oft nicht linear (Beispiel: Sensitivität der PD auf den ROA nimmt bei grossen Werten ab) und auch nicht immer monoton ist (Beispiel: sowohl ein starker Anstieg als auch ein starker Rückgang des Umsatzes kann zu erhöhter Ausfallwahrscheinlichkeit führen). Gleichzeitig ist ein zentrales Element des Modells seine Transparenz, d.h. dass es für den Nutzer intuitiv sein muss, warum die Änderung einer Inputvariablen zu einer Verände-

rung der PD führt. Um die Anforderung an Transparenz und die Berücksichtigung nicht-linearer Zusammenhänge zu erfüllen, basiert das Modell auf einer zweistufigen Funktion mit nicht-parametrischen Transformationen. Es gehört damit zu einer Klasse von Modellen, die als ‚Generalized Additive Models‘ bezeichnet werden (dazu auch Hastie/Tibshirani, 1990).

II. Datenbasis

Wir analysieren die Risikoklassifizierung der drei Modelltypen anhand eines realen KMU-Kreditportfolios einer schwergewichtig im Bilanzgeschäft tätigen und regional ausgerichteten Bank. Das verwendete Kreditportfolio umfasst 435 Gegenparteien. Die Kreditnehmer sind Schweizer KMU, deren Kreditexposure mindestens CHF 100'000 und deren Bruttoumsatz mindestens CHF 1.5 Mio. und höchstens CHF 100 Mio. beträgt. Die Abbildung 1 zeigt die Grössenverteilung im Kreditportfolio nach Umsatz.

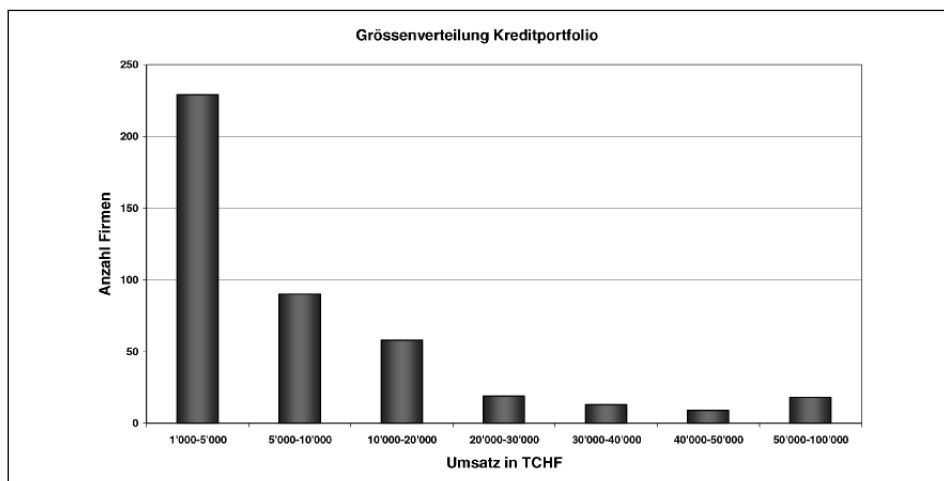


Abb. 1. Grössenverteilung Kreditportfolio

Es handelt sich dabei um das Gesamtportfolio der Bank, bereinigt um einige wenige Kreditkontrakte, für die aus unbekannten Gründen einzelne Datenelemente nicht erfasst wurden. Für jeden Kunden liegen zu dem entsprechenden Rating detaillierte Daten aus mehreren aufeinander folgenden Jahresabschlüssen (Bilanz und Gewinn- und Verlustrechnung) sowie zahlreiche qualitative Daten vor. Dabei definieren wir quantitative Daten als Merkmale, die auf publizierten Jahresabschlussdaten basieren wie z.B. Bruttoertrag, kurz- und langfristige Verbindlichkeiten, Anteil Fremdkapital oder Firmenalter. Unter qualitativen Daten verstehen wir Daten, deren Erhebung teilweise oder vollständig auf der Bewertung nicht publizierter und subjektiver Attribute basiert. Beispiele dazu sind die Markt- und Konkurrenzsituation des Unternehmens, die Beurteilung der Fähigkeit des Managements oder der Nachfolgeplanung. Die qualitativen Daten wurden durch die

Kreditanalysten der Bank erhoben beziehungsweise resultieren auf deren Einschätzung. Für alle Elemente des Samples standen Stichtagswerte des letzten Jahresabschlusses sowie die Vorjahresdaten zur Verfügung.

Die Unternehmen des Samples sind gut diversifiziert in Bezug auf das Firmenalter, das von Neugründungen mit einem Alter <1 Jahr bis zu Firmen mit einem Alter von 136 Jahren bei einem Mittelwert von 22.52 Jahren reicht. Die Industrieverteilung des Portfolios ist in der Abbildung 2 zusammengefasst (Immobilienunternehmen wurden nicht berücksichtigt):

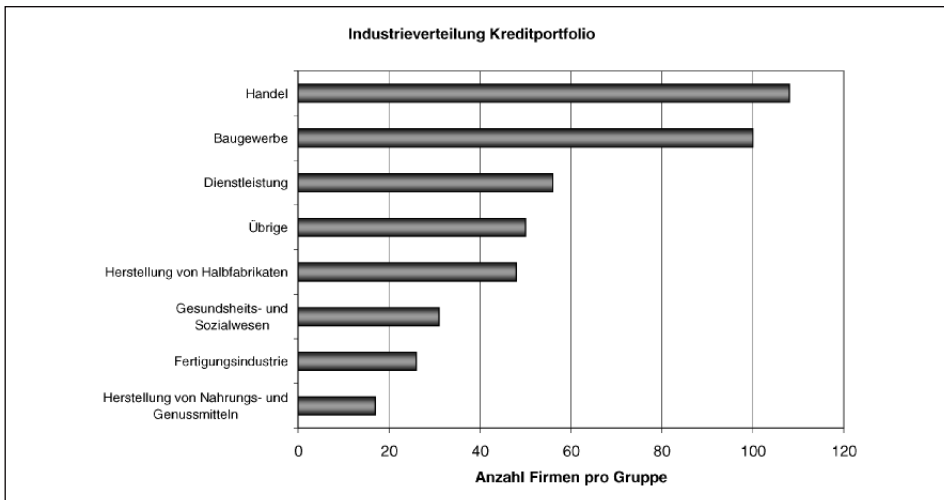


Abb. 2. Industrieverteilung Kreditportfolio

Das Kreditportfolio ist damit hinsichtlich Grössen- und Industrieverteilung gestützt auf Kley (2003) für den Schweizer KMU-Markt als repräsentativ zu bezeichnen. Die im Datenset vorhandene Konzentration auf eine Region hat weder Einfluss auf die Branchenverteilung des Portfolios noch auf die nachfolgende Bonitätseinschätzung.

III. Vorgehen

Für die im Datenset enthaltenen Unternehmungen wurden für alle drei Modelle Ratingklassifikationen generiert. Es wird ein identischer Ratingstichtag unterstellt. Zuerst wurden die Ratingklassifikationen anhand des Modells 1 erhoben. Kunden, die aufgrund des Ratings mit diesem Modell als ‚impaired‘ klassifiziert sind, wurden aus dem Sample entfernt, da die Defaultdefinitionen der weiteren Ratingmodelle nicht identisch mit derjenigen aus Modell 1 sind. Die quantitativen Daten wurden für die Bewertung mit den drei Ratingmodellen so aufbereitet, dass sie der jeweiligen Bilanzierungslogik entsprechen, die Daten richtig eingelesen und die benötigten Kennzahlen berechnet werden konnten. Die vorliegenden qualitativen Daten waren ausreichend für die Bewertung mit den Modellen 1 und 2. Es war lediglich ein Mapping auf die jeweiligen Skalen zur Be-

wertung einer Ausprägung notwendig. Für Modell 3 war die Erfassung der qualitativen Daten nicht erforderlich (reines Bilanzdatenmodell). Hier wurde für die Ratingermittlung jedoch zusätzlich mit Werten aus der Vorjahresbilanz gearbeitet.

Die Ergebnisse der Bewertungen wurden in der Logik der jeweiligen Ratingskala mit entsprechendem 1-Jahres-PD geliefert. Anschliessend wurde zur Sicherstellung der Vergleichbarkeit der Ergebnisse über die PD ein Mapping auf die offizielle Ratingskala von Moody's vorgenommen, für welche die PD-Bandbreiten der jeweiligen Ratingklassen bekannt sind. Fiel die PD auf eine Grenze der Moody's PD-Bandbreiten, wurde die jeweils schlechtere Ratingklasse gewählt.

C. Zusammenfassung der Ergebnisse

- I. H_1 : Unterschiedliche Ratingarchitektur führt für den Kreditnehmer c.p. zu einem signifikant unterschiedlichen Rating

Um die Hypothese H_1 zu überprüfen, untersuchen wir die Ratingergebnisse der drei Modelle im Vergleich. Dabei verwenden wir stets die mit Hilfe der originären PDs auf die Ratingskala von Moody's gemappten Ratingergebnisse unter Verwendung der entsprechenden ordinalen Klassifizierungen nach Moody's.⁴ Die Analysen werden zunächst mit Hilfe von beschreibender Statistik durchgeführt und anschliessend durch formale statistische Tests (Test auf Korrelation/Ähnlichkeit mit Bravais-Pearson/Spearman, Test auf identische Verteilung mit Wilcoxon Rangsummentest) überprüft.

Abbildung 3 zeigt die Verteilungen der Ratingausprägungen in der Stichprobe von 435 Unternehmen für alle drei Modelle im Vergleich.

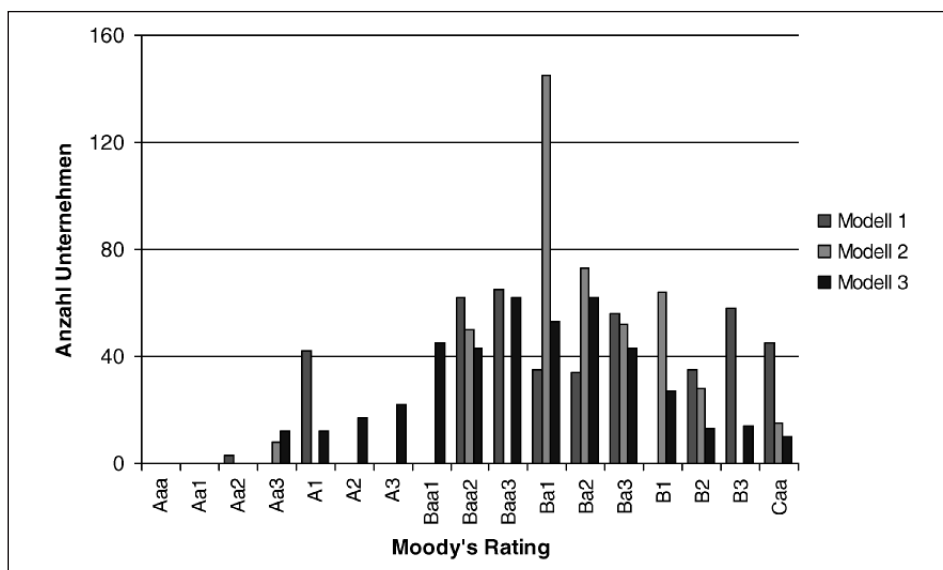


Abb. 3. Histogramm Rating

Bereits die Häufigkeitsverteilung der Ratingklassifikationen zeigt für die drei Modelle ein deutlich unterschiedliches Bild. Sind die Werte bei Modell 1 und Modell 3 breit über die Ratingskala gestreut, zeigt sich bei Modell 2 eine starke Konzentration von Werten im Bereich von Ba1 und schlechteren Ratingklassen.

Abbildung 4 zeigt anhand einer Boxplot-Darstellung die Ratingausprägungen der drei Modelle im Vergleich. Die grösste Streuung der Werte zwischen dem 25% und dem 75% Quantil findet sich beim Modell 1; Modell 2 weist eine mittlere und Modell 3 die geringste Streuungsbandbreite auf. Diese Darstellung stützt die Erkenntnisse aus der Analyse der Histogramme in Abbildung 3, erlaubt darüber hinaus jedoch die Bildung einer Reihenfolge der Varianz der Ratingergebnisse. Obwohl sowohl Modell 1 als auch Modell 3 die Breite der Ratingskala nutzen, sind die Werte bei Modell 1 gleichmässiger über alle Klassen verteilt, was zu einer grösseren Varianz führt.⁵

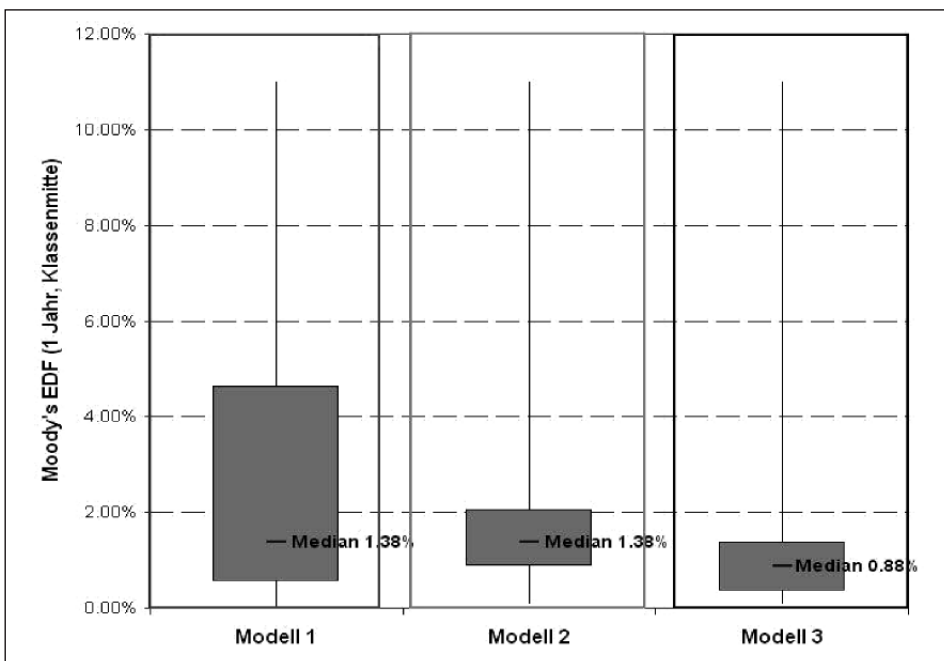


Abb. 4. Boxplots Ratingmodelle

Auch die Analyse der Mediane der drei Modelle gibt aus einer Portfoliooptik Hinweise auf unterschiedliche Ratingergebnisse der Modelle trotz identischem Dateninput. Die Mediane von Modell 1 und 2 sind identisch; sie weisen auf eine höhere Merkmalsausprägung (schlechtere Bonität) als Modell 3 hin. Die wertmässige Überlappung des Boxplots von Modell 3 mit den Boxplots zu Modell 1 und 2 ist ausserdem relativ gering, da dieser mehrheitlich unterhalb der anderen Boxplots liegt. Dies ist ein Indiz für die Unterschiedlichkeit der Verteilungen, das im Folgenden anhand der Cumulative Distribution Function jedes Ratingmodells näher überprüft wird. Abbildung 5 zeigt die Ergebnisse für alle Ratingmodelle im Vergleich.

Die graphische Darstellung der kumulativen Verteilungen legt zunächst einmal nahe, dass die Daten korreliert sind. Zur Überprüfung werden die Korrelation bzw. Ähnlichkeitsmasse nach Bravais-Pearson (unter der Annahme, dass die Daten bei über 400 Beobachtungen näherungsweise metrisch sind) und nach Spearman (nicht-parametrisch) berechnet. Die Korrelationen sind in beiden Fällen signifikant auf dem 0.01-Level.

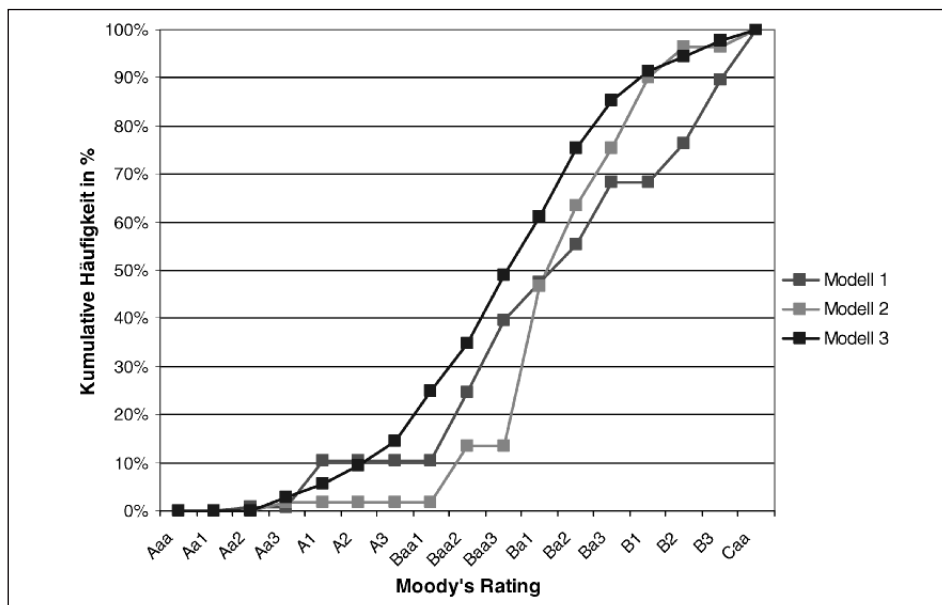


Abb. 5. Kumulative Verteilungsfunktion

Weiterhin unterstützt die Graphik die Vermutung, dass die empirischen Ergebnisse der drei Modelle nicht von derselben Verteilung stammen. Die Werte von Modell 1 und 2 liegen konstant unter den Werten von Modell 3 (mit Ausnahme der Verteilungsränder). Zusätzlich ändert sich das Verhalten von Modell 1 und 2 mit schlechter werdenden Ratings: zunächst steigen die Werte für Modell 1 im guten Ratingbereich schneller an, ab Ba1 verkleinert sich die Steigung jedoch, die Kurven schneiden sich ungefähr im Median und danach steigen die Werte von Modell 2 stärker an. Das zeigt, dass Modell 2 höhere kumulative Häufigkeiten im schlechten Ratingbereich aufweist.

Um statistisch signifikante Aussagen über die den Ratingergebnissen zugrunde liegenden Verteilungen machen zu können, vergleichen wir in einem nächsten Schritt die Stichprobenwerte paarweise mit Hilfe eines Wilcoxon-Rangsummen-Tests (WRS-Test). Dieser Test ist ein nicht-parametrischer statistischer Test, der sich auf abhängige Samples wie dem vorliegenden anwenden lässt. Wir testen die Nullhypothese, dass je zwei Samples von derselben gemeinsamen, unbekannten Verteilung stammen. Die Nullhypothese kann in allen drei Paarvergleichen auf Basis eines Signifikanzniveaus von 0.01 abgelehnt werden. Die Verteilungen der Ratingergebnisse der drei untersuchten Modelle stammen nicht von derselben Verteilung, auch wenn sie positiv korreliert sind.

Zusammenfassend bestätigen diese statistischen Auswertungen in eindrücklicher Weise die Hypothese, dass identische Risikoexposures in Abhängigkeit von der Architektur des Ratingmodells zu substantiell unterschiedlichen Risikobewertungen führen können. Zwar gingen Theorie wie Bankpraxis schon bisher implizit von dieser Vermutung aus; mit Hilfe unseres für einen breiten (schweizerischen) KMU Markt repräsentativen Datensamples sowie der drei eingesetzten typischen Ratingmodelle kann diese Hypothese jedoch erstmals auch empirisch bestätigt werden.

II. H_2 : Der Einbezug qualitativer Informationen führt zu einer signifikant höheren Bonitätseinschätzung im Vergleich zur Ratingevaluation auf der Basis rein quantitativer Informationen

In einem nächsten Schritt untersuchen wir, ob und in welchem Ausmass der Einbezug qualitativer Informationen c.p. das Ratingergebnis bzw. die Risikoklassifikation beeinflusst. Zur Überprüfung der Hypothese 2 vergleichen wir die drei Modelle paarweise. Wir evaluieren die Übereinstimmung der Ratingergebnisse je Ratingklasse und stellen für jede Ratingklasse die Abweichungen dar, indem wir untersuchen, wie die einzelnen Kreditkontrakte, die von einem Modell einer bestimmten Ratingklasse zugeordnet werden, durch das andere Modell klassifiziert werden. Indem wir die Abweichungen in Bezug zu den unterschiedlichen Architekturen der Modelle setzen, können wir Rückschlüsse über die Bedeutung der qualitativen und quantitativen Faktoren für die Ratingklassifizierung ziehen.⁶

1. Gegenüberstellung Modell 1/Modell 2

Nachfolgend die Gegenüberstellung der Ratingergebnisse aus Modell 1 bzw. Modell 2:

Tab. 1. Gegenüberstellung Modell 1/Modell 2

| Modell 2 | Modell 1 | | | | | | | | | | | | |
|----------|----------|-----|----|------|------|-----|-----|-----|----|----|----|-----|-----|
| | Aa2 | Aa3 | A1 | Baa2 | Baa3 | Ba1 | Ba2 | Ba3 | B1 | B2 | B3 | Caa | |
| Aa2 | | | | | | | | | | | | | 0 |
| Aa3 | | | | | 1 | | | 7 | | | | | 8 |
| A1 | | | | | | | | | | | | | 0 |
| Baa2 | 1 | | 17 | 10 | 11 | 3 | 4 | | | 3 | | 1 | 50 |
| Baa3 | | | | | | | | | | | | | 0 |
| Ba1 | 2 | | 18 | 39 | 29 | 14 | 15 | 7 | | 11 | 7 | 3 | 145 |
| Ba2 | | | 3 | 8 | 12 | 6 | 5 | 14 | | 7 | 13 | 5 | 73 |
| Ba3 | | | 3 | 1 | 4 | 4 | 5 | 11 | | 6 | 12 | 6 | 52 |
| B1 | | | 1 | 4 | 4 | 6 | 4 | 8 | | 5 | 14 | 18 | 64 |
| B2 | | | | | 3 | 2 | 1 | 7 | | 2 | 4 | 9 | 28 |
| B3 | | | | | | | | | | | | | 0 |
| Caa | | | | | 1 | | | 2 | | 1 | 8 | 3 | 15 |
| | 3 | 0 | 42 | 62 | 65 | 35 | 34 | 56 | 0 | 35 | 58 | 45 | 435 |

Die beiden Modelle unterscheiden sich, wie in Abschnitt B.I gezeigt, in erster Linie durch den Einbezug bzw. die Gewichtung der qualitativen Faktoren. Die Gegenüberstellung des banktypischen Modells 1 und des hinsichtlich Kriterieneinsatzes komplexeren Modells 2 zeigt, dass ‘bessere’ Ratingergebnisse des Modells 1 (Ratingklassen Aa2-Ba1) im Modell 2 fast immer schlechter gerated werden, während die durch das Modell 1 tiefer gerateten KMU (Ratingklassen Ba2-Caa) im Modell 2 überwiegend höher eingestuft werden. Umgekehrt werden Kontrakte, die vom Modell 2 der oberen Hälfte der Ratingskala zugeordnet werden, im Modell 1 tendenziell leicht höher bewertet, diejenigen der unteren Hälfte der Ratingskala dagegen leicht tiefer:

Tab. 2. Zusammenfassung Gegenüberstellung Modell 1/Modell 2

| Durch Modell 2 | Bewertung durch Modell 1 | | Durch Modell 1 | Bewertung durch Modell 2 | |
|------------------------|---------------------------------------|---------------------------------------|------------------------|---------------------------------------|---------------------------------------|
| | 1. Hälfte Ratingskala (Aa2-Ba1) | 2. Hälfte Ratingskala (Ba2-Caa) | | 1. Hälfte Ratingskala (Aa2-Ba1) | 2. Hälfte Ratingskala (Ba2-Caa) |
| besser bewertet | 15 | 171 | besser bewertet | 135 | 69 |
| schlechter bewertet | 168 | 36 | schlechter bewertet | 112 | 74 |
| gleich bewertet | 24 | 21 | gleich bewertet | 24 | 21 |

Modell 2 gewichtet (wie in Abschnitt B.I beschrieben) bei tieferen Ratingklassifizierungen auf Basis des quantitativen Ratings die qualitativen Faktoren immer schwächer. In den untersten Ratingkategorien werden überhaupt nur noch quantitative Faktoren zur Risikoklassifizierung herangezogen. Modell 1 dagegen belässt die Gewichtung der qualitativen Faktoren über die ganze Bandbreite der Ratingskala hinweg unverändert. Offensichtlich führt die geringere Gewichtung der qualitativen Faktoren bei tieferen Ratingklassen im Modell 2 nicht zum aufgrund bisheriger empirischer Arbeiten erwarteten Ergebnis einer im Vergleich zum Modell 1 schlechteren Risikoklassifizierung (Brunner/Krahn/Weber, 2000).

Die Hypothese H₂ kann damit aufgrund der Gegenüberstellung von Modell 1 und Modell 2 nicht bestätigt werden. Der Einbezug qualitativer Faktoren in die Bonitätsklassifizierung führt nicht zu systematisch besseren Ergebnissen.

2. Gegenüberstellung Modell 1/Modell 3

Der Vergleich der Klassifizierungsergebnisse des hybriden Modells 1 mit dem ausschliesslich auf quantitativen Daten basierenden Modell 3 in Tabelle 2 zeigt ein ähnliches Bild.

Tab. 3. Gegenüberstellung Modell 1/Modell 3

| | Modell 1 | | | | | | | | | | | | | | | | |
|----------|----------|-----|----|----|----|------|------|------|-----|-----|-----|----|----|----|-----|-----|--|
| Modell 3 | Aa1 | Aa3 | A1 | A2 | A3 | Baa1 | Baa2 | Baa3 | Ba1 | Ba2 | Ba3 | B1 | B2 | B3 | Caa | | |
| Aa1 | | | | | | | | | | | | | | | | 0 | |
| Aa3 | 1 | 9 | | | | 2 | | | | | | | | | | 12 | |
| A1 | 1 | 7 | | | | 2 | | | | 1 | | 1 | | | | 12 | |
| A2 | | 4 | | | | 7 | 4 | 1 | 1 | | | | | | | 17 | |
| A3 | 1 | 5 | | | | 2 | 5 | 2 | 1 | 5 | | | | 1 | | 22 | |
| Baa1 | | 4 | | | | 15 | 14 | 7 | 1 | 3 | | 1 | | | | 45 | |
| Baa2 | | 6 | | | | 11 | 12 | 3 | 4 | | | 4 | 3 | | | 43 | |
| Baa3 | | 5 | | | | 18 | 10 | 7 | 9 | 6 | | 4 | | 2 | | 61 | |
| Ba1 | | 1 | | | | 3 | 10 | 5 | 8 | 11 | | 5 | 6 | 4 | | 53 | |
| Ba2 | | | | | | 1 | 5 | 8 | 7 | 13 | | 10 | 15 | 4 | | 63 | |
| Ba3 | | | | | | 1 | 5 | 2 | 2 | 7 | | 9 | 11 | 6 | | 43 | |
| B1 | | | | | | | | | | 4 | | | 11 | 12 | | 27 | |
| B2 | | 1 | | | | | | | 1 | | | | 6 | 5 | | 13 | |
| B3 | | | | | | | | | | 5 | | 1 | 3 | 5 | | 14 | |
| Caa | | | | | | | | | | 1 | | | 2 | 7 | | 10 | |
| | 3 | 42 | 0 | 0 | 0 | 62 | 65 | 35 | 34 | 56 | 0 | 35 | 58 | 45 | 0 | 435 | |

Die durch das Modell 1 positiv gerateten Unternehmungen (Aa1-Baa3) erhalten im rein quantitativ ausgerichteten Modell 3 überwiegend schlechtere Klassifizierungen zugewiesen. Umgekehrt bewertet Modell 3 die vom Modell 1 der unteren Hälfte der Ratingskala zugeordneten Kontrakte signifikant besser. Geht man vom Ratingergebnis des 3. Modells aus, so zeigt sich ein ähnliches Bild. Bewertungen des Modells 3 werden durch das Modell 1 tendenziell nach unten korrigiert, d.h. Unternehmungen erhalten im Modell 1 tendenziell über die ganze Ratingskala des Modells 3 hinweg schlechtere Bewertungen. Dieses Ergebnis verwirft wiederum die Hypothese H₂, nach der der Einbezug qualitativer Informationen tendenziell zu einer höheren Bonitätseinschätzung führt als bei ausschliesslicher Berücksichtigung quantitativer Informationen.

Tab. 4. Zusammenfassung Gegenüberstellung Modell 1/Modell 3

| Durch Modell 3 | Bewertung durch Modell 1 | | Durch Modell 1 | Bewertung durch Modell 3 | |
|---------------------|----------------------------------|---------------------------------|---------------------|----------------------------------|---------------------------------|
| | 1. Hälfte Ratingskala (Aa1-Baa3) | 2. Hälfte Ratingskala (Ba1-Caa) | | 1. Hälfte Ratingskala (Aa1-Baa3) | 2. Hälfte Ratingskala (Ba1-Caa) |
| besser bewertet | 49 | 156 | besser bewertet | 72 | 83 |
| schlechter bewertet | 115 | 40 | schlechter bewertet | 96 | 109 |
| gleich bewertet | 43 | 32 | gleich bewertet | 43 | 32 |

3. Gegenüberstellung Modell 2/Modell 3

Schliesslich vergleichen wir die Ratingergebnisse aus unserem Sample-Portfolio auch noch an Hand der Gegenüberstellung der Modelle 2 und 3. Wir vergleichen hier also ein hybrides Modell, das die Gewichtung qualitativer Faktoren in Relation zur Ratingklassifizierung variiert (Modell 2), mit einem rein quantitativ ausgerichteten Modell (Modell 3).

Tab. 5. Gegenüberstellung Modell 2/Modell 3

| Modell 3 | Modell 2 | | | | | | | | | | | | | | |
|----------|----------|----|----|----|------|------|------|-----|-----|-----|----|----|----|-----|-----|
| | Aa3 | A1 | A2 | A3 | Baa1 | Baa2 | Baa3 | Ba1 | Ba2 | Ba3 | B1 | B2 | B3 | Caa | |
| Aa3 | | | | | | 3 | | 8 | | 1 | | | | | 12 |
| A1 | 1 | | | | | 4 | | 5 | 2 | | | | | | 12 |
| A2 | | | | | | 5 | | 10 | 2 | | | | | | 17 |
| A3 | 4 | | | | | 7 | | 8 | 3 | | | | | | 22 |
| Baa1 | | | | | | 11 | | 29 | 4 | | | | | | 44 |
| Baa2 | | | | | | 10 | | 20 | 8 | 2 | 2 | | | 1 | 43 |
| Baa3 | 1 | | | | | 5 | | 31 | 16 | 4 | 3 | 2 | | | 62 |
| Ba1 | 2 | | | | | | | 13 | 10 | 10 | 13 | 3 | | 2 | 53 |
| Ba2 | | | | | | 3 | | 17 | 12 | 12 | 13 | 4 | | 2 | 63 |
| Ba3 | | | | | | 2 | | 4 | 11 | 12 | 8 | 5 | | 1 | 43 |
| B1 | | | | | | | | | 5 | 5 | 12 | 3 | | 2 | 27 |
| B2 | | | | | | | | | | 5 | 6 | 1 | | 1 | 13 |
| B3 | | | | | | | | | | 1 | 3 | 7 | | 3 | 14 |
| Caa | | | | | | | | | | | 4 | 3 | | 3 | 10 |
| | 8 | 0 | 0 | 0 | 0 | 50 | 0 | 145 | 73 | 52 | 64 | 28 | 0 | 15 | 435 |

Die Gegenüberstellung zeigt auch hier ein die Hypothese H_2 widerlegendes Bild. Über fast alle Ratingklassen hinweg resultiert für das untersuchte Sample aus dem Modell 2 trotz Einbezug qualitativer Faktoren (zumindest bis zur Stufe Baa2) ein teils deutlich schlechteres Rating als im Modell 3. Einzig die tieferen Ratingklassen ab Ba3 scheinen die Hypothese H_2 zu bestätigen, obgleich gerade in diesen Klassen auch im Modell 2 der Anteil der qualitativen Faktoren im Modell gegen Null tendiert. Wir müssen also auch auf der Basis dieser Gegenüberstellung die Hypothese H_2 verwerfen.

Tab. 6. Zusammenfassung Gegenüberstellung Modell 2/Modell 3

| Durch Modell 3 | Bewertung durch Modell 2 | | Durch Modell 2 | Bewertung durch Modell 3 | |
|--------------------------|--|---------------------------------------|--------------------------|--|---------------------------------------|
| | 1. Hälfte Ratingskala (Aa3-Baa3) | 2. Hälfte Ratingskala (Ba1-Caa) | | 1. Hälfte Ratingskala (Aa2-Baa3) | 2. Hälfte Ratingskala (Ba1-Caa) |
| besser bewertet | 30 | 253 | besser bewertet | 11 | 78 |
| schlechter bewertet | 18 | 71 | schlechter bewertet | 191 | 92 |
| gleich bewertet | 10 | 53 | gleich bewertet | 10 | 53 |

D. Interpretation und Schlussfolgerungen

Unsere Untersuchung prüft und bestätigt erstmals empirisch auf der Grundlage eines für den schweizerischen KMU-Markt repräsentativen Datenpools die in Theorie und Praxis seit langem geäußerte Vermutung, dass die Architektur eines Ratingmodells ein wichtiger Bestimmungsfaktor der Risikoklassifizierung eines Kredits ist. Der Vergleich der Ratingergebnisse der drei Modelle zeigt eine überraschend grosse Abweichung der Risikoklassifizierungen trotz identischem Dateninput. Nicht einmal die Hälfte (38-45%) der Ratingergebnisse sind im Rahmen der paarweisen Gegenüberstellung identisch oder liegen in einer Bandbreite von plus/minus einer Ratingklasse. Unsere Untersuchung bestätigt damit die Analyse von Jacobson/Lindé/Roszbach (2006), die anhand eines schwedischen Datensamples zeigen, dass die Quantifizierung des Kreditrisikos nicht nur vom formalen Design eines Ratingmodells abhängt (beispielsweise der Zahl der Ratingklassen), sondern dass eine Vielzahl weiterer Parameter das Ratingergebnis beeinflusst. Eine Untersuchung von Carey (2002), die in einem ähnlichen Ansatz wie dem hier gewählten verschiedene interne Ratingmodelle von Banken gegenüberstellt und auf einem Datensatz von 20 US-Banken basiert, zeigt auf, dass die Modelle in 45% der Fälle zu einem vergleichbaren Ergebnis kommen und in gut 95% der Fälle die Abweichungen innerhalb einer Bandbreite von 2 Notches liegen. Die Hypothese H_1 , wonach unterschiedliche Ratingarchitekturen zu signifikant unterschiedlichen Ratingergebnissen führen, konnte also klar bestätigt werden. Für dieses Ergebnis können verschiedene Faktoren verantwortlich sein. Grundsätzlich verfolgen zwar alle drei Modelle dasselbe Ziel und messen identische Risiken. Sie unterscheiden sich aber hinsichtlich der jeweiligen Gewichtung der qualitativen bzw. quantitativen Faktoren, der Gewichtung der jeweiligen Faktorgruppen (Summe der quantitativen bzw. qualitativen Faktoren), der Kalibrierung in Bezug auf die Ausfallraten, der zugrunde liegenden Ausfalldefinition der Daten, der Transformation der Ausprägungen in Scores usw.

Wir können zudem anhand der drei typischen Architekturen der von uns eingesetzten Ratingmodelle erstmals zeigen, dass aus unterschiedlichem Architekturdesign nicht nur eine andere Ratingklassifizierung resultiert, sondern dass sich auch die Ratingverteilungsfunktionen eines KMU-Portfolios verändern. Wie sich aus der Beschreibung der Modellarchitekturen in Abschnitt B.I ableiten lässt, unterscheiden sich die drei für die Analyse

gewählten Modelle auch hinsichtlich Trennschärfe der Ratingklassen, was c.p. natürlich unabhängig von Einbezug und Gewichtung qualitativer Faktoren zu einer unterschiedlichen Verteilung über die Bonitätsstufen führt. Diese Verteilung der Risikoexposures in einem Kreditportfolio beeinflusst wichtige Parameter der bankbetrieblichen Kreditpolitik wie etwa die Bildung von Rückstellungen, die Ausgestaltung von Covenants oder die Bestimmung von Limiten. Über die Berechnung von in den Kreditpreis einflussenden Risikoprämien wird eine Verteilungsfunktion indirekt auch relevant für die wettbewerbsstrategische Positionierung eines Kreditanbieters im Markt.

Diese Erkenntnisse können sich aber auch auf das Verhalten des Kreditnachfragers auswirken. Kreditnachfrager, die sich der Konsequenzen unterschiedlicher Modellarchitekturen bewusst sind, werden verstärkt Ratingarbitrage betreiben. Die Entstehung unabhängiger KMU-Ratingagenturen und die Akzeptanz bankexterner Ratings durch die Aufsichtsbehörden werden diese Entwicklung unterstützen. Für die kreditgebende Bank entsteht damit ein Problem der adversen Selektion. Aus einer makroökonomischen Perspektive schliesslich stellt sich die Frage, wie effizient knappe Kapitalressourcen risikobehafteten Projekten zugeordnet werden. Auch Fender/Kiff (2005) kommen in einer vergleichbaren Studie, die sich allerdings auf den Vergleich von Ratingmodellen für Verbriefungstransaktionen bezieht, zu dem Ergebnis, dass bei gleichem Input und unterschiedlichen Ratingergebnissen aufgrund von modellarchitektonischen Unterschieden die Gefahr von Arbitrage besteht.

Unsere empirische Analyse zeigt ein weiteres interessantes Ergebnis. Der Einbezug qualitativer Faktoren führt zwar teilweise, nicht jedoch systematisch zu einer Ratingverbesserung im Sinne einer erhöhten Bonitätseinschätzung beim Gesamtrating. Dieses Erkenntnis relativiert Aussagen der bisher zu dieser Thematik veröffentlichten Literatur (vgl. etwa Brunner/Krahn/Weber, 2000; Grunert/Norden/Weber, 2005). Ursache für diese Diskrepanz könnte sein, dass die Primäreinteilung aufgrund der quantitativen Kriterien durch Modell 2 sehr restriktiv erfolgt (siehe dazu die Charakterisierung der Ratingmodelle in Kapitel B.I). Die qualitativen Ratings können in der anschliessenden Bewertung zwar einen positiven Einfluss auf das Gesamtrating ausüben, dieser reicht jedoch nicht aus, um die schlechtere Primäreinteilung im Vergleich zu Modell 3 auszugleichen. Dem Design nach kann durch das quantitative Rating maximal eine Klasse 5 von 10 erreicht werden; durch die qualitativen Faktoren kann eine maximale Ratingverbesserung auf Ratingstufe 3 erfolgen. So kann kein gesamthaft positiveres Ergebnis als bei den rein auf quantitativen Kriterien beruhenden Ratingmodellen resultieren, obwohl die qualitativen Faktoren einen positiven Einfluss haben könnten.

Bei der Beurteilung unserer Analyseergebnisse gilt es zu beachten, dass wir die Ratingmodelle – mit Ausnahme der Beschreibung der Basisarchitektur – als ‚Black Box‘ betrachtet haben, so dass eine direkte Korrelation der Ratingergebnisse mit einzelnen Faktorgruppen (quantitativ/qualitativ) nur eingeschränkt auf die Modellarchitektur zurückgeführt werden kann. Eine weitere, mit dem vorliegenden Untersuchungsdesign nicht weiter analysierbare Erklärung wäre eine in der Modellarchitektur eingebaute Veränderung von Gewichtungsfaktoren in Abhängigkeit der Ratingklasse (wie beispielsweise das Überschreiten von Grenzwerten bei Liquiditäts- oder Eigenkapitalratios, das im Modell 1 zu einer überproportionalen Verschlechterung der Ratingklassifikation führen kann). Als dritte Erklärung wäre eine positive Korrelation quantitativer und qualitativer

Faktoren denkbar, indem beispielsweise schlechte quantitative Werte automatisch zu einer schlechteren Beurteilung der qualitativen Faktoren führen und damit das ‚Downgrading‘ bei Model 1 überproportional verstärken (d.h. gute Ratingklassifizierungen werden systematisch zu gut, schlechte zu gering bewertet).

E. Zusammenfassung/Ausblick

Wir haben anhand eines für den schweizerischen KMU-Markt repräsentativen Portfolios insgesamt 435 Unternehmungen mit drei verschiedenen Ratingmodellen bewertet. Die untersuchten Modelle verfolgen alle denselben Zweck der Bonitätseinschätzung von KMU-Gegenparteien. Aufgrund der unterschiedlichen Architektur, der Kalibrierung der Modelle und der unterschiedlichen Gewichtung qualitativer Kriterien resultieren teilweise substantielle Unterschiede in der Risikoklassifizierung bzw. im Rating. Die Hypothese H_1 , dass die Ratingarchitektur eine wesentliche Determinante des Ratingergebnisses ist, kann somit überzeugend bestätigt werden. Diese Erkenntnis hat Auswirkungen auf wichtige Parameter der bankbetrieblichen Kreditpolitik wie auch auf das Verhalten der Kreditnachfrager.

Die drei getesteten Modelle unterscheiden sich insbesondere hinsichtlich der Gewichtung qualitativer Faktoren im Rahmen des Ratings. Die paarweise Gegenüberstellung der Ratingergebnisse der Modelle zeigt, dass der Einbezug qualitativer Faktoren nicht zu einer systematischen Höherbewertung der Bonität des Kreditnehmers führt. Die Hypothese H_2 muss deshalb verworfen werden.

Im Rahmen weitergehender statistischer Analysen wird der Einfluss der qualitativen Kriterien auf das Ratingergebnis von statistischen Ratingtools zu untersuchen sein. Dazu sind in erster Linie detailliertere Datenpools als die bisher zur Verfügung stehenden nötig, in denen beispielsweise Ausfalldaten, Daten zu den Ratingergebnissen der Einzelanalysen aufgrund von quantitativen und qualitativen Faktoren oder Informationen zu den genauen Transformationsvorschriften in den Ratingmodellen enthalten sind. Mit dieser Informationsbasis liessen sich auch vertiefte Untersuchungen zur Validierung und Gütemessung unterschiedlicher Architekturen anstellen.

Anmerkungen

- 1 Siehe zu Details über ‚point-in-time‘ versus ‚through-the-cycle‘ Ratings u.a. Rösch (2005).
- 2 Siehe zu Details über den Scoringprozess auch Brunner/Krahen/Weber (2000), Altman (2000), Hastie/Tibshirani (1990).
- 3 Siehe generell zur Diskussion von Trennschärfemassen für Ratingmodell u.a. Stein (2005), Blochwitz et al. (2004).
- 4 Es wurde hier bewusst auf die Verwendung der originären PDs verzichtet, da sich bei der Analyse gezeigt hat, dass der Erkenntnisgewinn aus der Verwendung der Originaldaten marginal ist und die Darstellbarkeit der Ergebnisse unverhältnismässig leidet. Der Informationsverlust durch das Mapping auf eine gemeinsame Skala ist gering und beeinflusst die Interpretation der Ergebnisse aus dem empirischen Test nicht.
- 5 Die Analyse der Maximalwerte ist nicht relevant, da die PD für die schlechteste Ratingklasse auf 11% limitiert wurde.
- 6 Die Tabellen der paarweisen Vergleiche zeigen teilweise eine unterschiedliche Anzahl Ratingklassen, da Klassen, in denen sich weder im einen noch im anderen Modell Werte befanden, aus Gründen der Übersichtlichkeit gelöscht wurden.

Literatur

- Altman, E.I./Rijken, H.A. (2004): How rating agencies achieve rating stability. *Journal of Banking and Finance*, 28 (2004), S. 2679–2714.
- Altman, E.I. (2000): Predicting financial distress of companies: Revising the Z-Score and ZETA Models. New York University, New York.
- Basler Ausschuss (2006): Internationale Konvergenz der Kapitalmessung und Eigenkapitalanforderungen – Überarbeitete Rahmenvereinbarung; April 2006.
- Basler Ausschuss (2000a): Range of Practice on Banks' Internal Rating Systems; Working Paper, January 2000; www.bis.org.
- Bernet, B./Westerfeld, S. (2007): Konzepte des KMU Rating im Vergleich. In: Everling, O. (Hrsg.): *Certified Rating Analyst*. Oldenbourg, München 2007.
- Blochwitz, S./Hammerle, A./Hohl, S./Raumeier, R./Rösch, D. (2004): Was leisten Trennschärfemasse für Ratingmodelle. *Kreditwesen* 22/2004, S. 41–44.
- Brunner, A./Krahn, J.P./Weber, M. (2000): Information Production in Credit Relationships: On the Role of Internal Ratings in Commercial Banking. CFS Working Paper No. 2000/10.
- Cantor, R./Packer, F. (1997): Differences of opinion and selection bias in the credit rating industry. *Journal of Banking and Finance*, 21, S. 1395–1417.
- Carey, M. (2002): Some evidence on the consistency of banks' internal credit ratings. In: Ong, M.K. (Ed): *Credit Ratings: Methodologies, Rationale and Default Risk*, London 2002.
- Eigermann, J. (2002): *Quantitatives Credit-Rating unter Einbeziehung qualitativer Merkmale*. Sternfels, 2. Aufl. 2002.
- English, W.B./Nelson, W.R. (1998): Bank Risk Rating of Business Loans. Board of Governors of the Federal Reserve System, April 1998.
- Fender, I./Kiff, J. (2005): CDO rating methodology: some thoughts on model risk and its implications. *Journal of Credit Risk*, Volume 1/ Number 3, Summer 2005, S. 37–58.
- Füser, K. (2001): Mittelstandsrating mit Hilfe neuronaler Netzwerke. In: Everling, O. (Hrsg.): *Rating – Chance für den Mittelstand nach Basel II*, S. 363–386, Gabler, Wiesbaden 2001.
- Grunert, J./Norden, L./Weber, M. (2005): The role of non-financial factors in internal credit ratings. *Journal of Banking and Finance*, Vol. 29 (2005), S. 509–531.
- Hastie, T.J./Tibshirani, R.J. (1990): *Generalized Additive Models*. Chapman and Hall, New York.
- Hornik, K./Jankowitsch, R./Lingo, M./Pichler, S. and Winkler, G. (2005). *Validation of Credit Rating Systems Using Multi-Rater Information*. Working Paper.
- Jacobson, T./Lindé, J./Roszbach, K. (2006): Internal ratings systems, implied credit risk and the consistency of banks' risk classification policies. *Journal of Banking and Finance* 30(2006), S. 1899–1926.
- Kley, C. (2003): *Mittelstands-Rating*. Deutscher Universitäts-Verlag, Wiesbaden 2003.
- Krahn, J./Weber, M. (2001): Generally Accepted Rating Principles: A Primer. *Journal of Banking and Finance*, Vol. 25 (2001), S. 3–23.
- Löffler, G. (2004): An Anatomy of Rating Through the Cycle. *Journal of Banking and Finance*, 28, S. 695–720.
- Mommel, C./Schmieder, C./Stein, I. (2006): Relationship Banking – Empirical Evidence from Germany. Working Paper Deutsche Bundesbank.
- Nakamura, L.I./Roszbach, K. (2005): Credit Ratings and Bank Monitoring Ability. Paper presented at the 2006 FMA Annual Meeting, October 2006; www.fma.org.
- Oelerich, A. (2004): Robuste Ratingverfahren. Zur Steigerung der Prognosequalität quantitativer Ratingverfahren. Wiesbaden 2004.
- Poon, W.P.H. (2003): Are unsliced credit ratings biased downward? *Journal of Economics and Business*, 27, S. 593–614.
- Rösch, D. (2005): An empirical Comparison of Default Risk Forecasts from Alternative Credit Rating Philosophies. *International Journal of Forecasting* 21 (2005), S. 37–51.
- Stein, R. (2005): The relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing. *Journal of Banking and Finance*, Vol 29 (2005), S. 1213–1235.
- Treacy, W.F./Carey, M. (2000): Credit risk rating systems at large US banks. *Journal of Banking and Finance*, 24 (2000), S. 167–201.
- Weber, M./Krahn, J.P./Vossmann, F. (1999): Risikomessung im Kreditgeschäft: Eine empirische Analyse bankinterner Ratingverfahren. *Zeitschrift für betriebswirtschaftliche Forschung, Sonderheft 41(1999)*, S. 117–142.

KMU-Ratingmodelle und Ratingqualität: Auswirkungen der Ratingarchitektur auf die ex-ante Risikoklassifikation von KMU-Kreditkontrakten

Zusammenfassung

Aus der Sicht einer KMU als aktuelle oder potentielle Kreditnachfragerin bei einer Bank stellt sich die Frage, ob eine unterschiedliche Ausgestaltung interner Ratingmodelle verschiedener Kreditanbieter bei grundsätzlich identischen Inputdaten zu unterschiedlichen Bonitätsklassifikationen und unterschiedlichen Kreditkonditionen führt. Mit dieser Fragestellung befasst sich der vorliegende Beitrag. Sie ist unter dem Aspekt optimaler Allokationseffizienz sowohl aus makroökonomischer wie aus mikroökonomischer Sicht von Bedeutung. Wir bewerten ein für den schweizerischen KMU-Markt repräsentatives Portfolio von insgesamt 435 KMU-Kreditnehmern mit drei verschiedenen Ratingmodellen, die typisch für die bei Banken zur Bewertung von KMU-Risiken eingesetzten Modelle stehen. Dabei testen wir zunächst die Frage, ob differierende Ratingarchitekturen für den Kreditnehmer c.p. zu einem signifikant unterschiedlichen Rating führen. Diese Vermutung kann aufgrund der Untersuchung bestätigt werden. Anschliessend untersuchen wir die Hypothese, ob der Einbezug qualitativer Informationen zu einer signifikant höheren Bonitätseinschätzung im Vergleich zur Ratingevaluation auf der Basis rein quantitativer Informationen führt. Diese Hypothese muss verworfen werden, was teilweise im Gegensatz zu bisherigen empirischen Erkenntnissen steht.

SME Rating Models and Rating Quality: How rating model architecture impacts ex-ante risk classification of SME credit contracts

Summary

Rating models used by banks to evaluate the credit worthiness of SME clients still differ substantially with regards to the underlying rating philosophy, system architecture and calibration. Looking at bank ratings from an SME perspective as a current or potential future borrower, the question arises, if – even though input data remains unchanged – the different designs of bank internal rating models lead to different rating results and subsequently to different credit conditions. This problem guides the research question at the heart of this article. Therefore, in our empirical study, we firstly rate a representative sample of Swiss SME clients by different rating models and test, if different rating models belonging to different types of model architectures lead to different risk classifications (rating results) even though identical input data is used. This first hypothesis can be supported, which could lead to rating model arbitrage against banks. The second hypothesis is based on empirical findings from other studies and analyses if the inclusion of qualitative information leads to significantly higher rating marks compared to ratings solely based on quantitative information. Our findings do not support the second assumption.